

Executive Steering
Committee
For A.C.E. Policy II
(ESCAP II)
Report 6

October 13, 2001

Census Person Duplication and the Corresponding A.C.E. Enumeration Status

Roxanne Feldpausch

Decennial Statistical
Studies Division

U S C E N S U S B U R E A U

Helping You Make Informed Decisions

CONTENTS

EXECUTIVE SUMMARY	iii
1. Background	1
2. Methods	2
3. Limits	2
4. Results	3
5. Summary	7
References	8
Appendix A	9
Appendix B: Technical Documentation	23
Appendix C: Assignment of the Unbiased Probability of Duplication	25

TABLES

Table 1 E-Sample Duplicates to People Outside the Surrounding Blocks	4
Table 2 E-Sample People Duplicated outside the A.C.E. search area to people in Group Quarters by type of Group Quarters	5
Table 3 E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters (Standard Errors)	6
Table 4.A E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters by PFU respondent	9
Table 4.B Percent Match, Correct Enumeration and Erroneous Enumeration of E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters by PFU Respondent	10
Table 5.A E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters by Age Group	11
Table 5.B Percent Match, Correct Enumeration and Erroneous Enumeration of E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters by Age Group ..	13
Table 6.A E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters by Census respondent	15
Table 6.B Percent Match, Correct Enumeration and Erroneous Enumeration of E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters by Census Respondent	16
Table 7.A E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters by Whole vs Partial HH Duplicate (Standard Errors)	17
Table 7.B Percent Match, Correct Enumeration and Erroneous Enumeration of E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters by Whole vs Partial HH Duplicate (Standard Errors)	18
Table 8.A E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters by Domain (Standard Errors)	19
Table 8.B Percent Match, Correct Enumeration and Erroneous Enumeration of E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters by Domain (Standard Errors)	21

EXECUTIVE SUMMARY

In past censuses, we have had no way to evaluate the A.C.E. coding of people duplicated outside the post-enumeration survey's search area. For Census 2000 there is an analysis project that searched for duplicates through-out the country. Matching work was conducted to determine the extent of duplicate enumerations that were not found in the A.C.E. See Mule (2001) for more details. This allowed us to evaluate the A.C.E. coding of people duplicated outside the search area.

A person found duplicated was captured twice by the census. The record of the person that was captured in the correct place according to census residence rules should have been coded as a correct enumeration. The record of the person that was captured in an incorrect place according to census residence rules should have been coded as an erroneous enumeration.

What is the scope of this report?

This analysis looks at how the A.C.E. coded E-Sample people duplicated to people outside the A.C.E. search area. We obtained large standard errors for many of the numbers of interest. This was due to the small number of cases considered in this study. This limited our ability to make meaningful interpretations.

What are the interesting findings?

The percent erroneous enumeration for E-Sample people duplicated to people in group quarters where the residents were not allowed to claim usual home elsewhere (45.5% for college dorms and 16.5% for other group quarters) was low. For the majority of these people, it seems likely that the groups quarters was the usual April 1 residence.

The percent erroneous enumeration for E-Sample people duplicated to people in group quarters where the residents were allowed to claim usual home elsewhere (12.5%) was high. For most of these people, it seems likely that the housing unit was the usual April 1 residence.

The percent erroneous enumeration for E-Sample people duplicated to people in housing units outside the A.C.E. search area (14.2%) was lower than the approximate fifty percent one might have expected. One might expect fifty percent because half of the time the wrong housing unit should be in sample, resulting in coding the residents as erroneous.

What are the possible explanations for these findings?

- The instructions indicating who to include on the census questionnaire may not have been completely understood or the instructions may have been understood, but were ignored by the respondent. Examples include college students, people in local jail and people in nursing homes.
- The respondent may have not realized that a household member was enumerated elsewhere.
- Some group quarters' enumeration may have been done using administrative records that did not reflect the residents as of April 1, 2000.
- Some group quarters are temporary, such as local jails. Some people counted here may actually be usual residents of the sample housing unit.
- The computer matching of duplicates outside the search area might be incorrect. We do not believe that this was a large part of the explanation, because we only looked at those cases that we had confidence in (those cases that had a high probability of being linked correctly).
- The A.C.E. did not do an optimal job of identifying people who should have been coded as erroneous enumerations due to other residence. The percent other residence was 1.4 in the 2000 A.C.E. and 2.3 in the 1990 PES (these percentages reflect the redistribution of people with unresolved status). The results of the Measurement Error Reinterview also measure this phenomenon.

What implications do these results have on the ESCAP adjustment decision?

There is evidence that the A.C.E. did not code as erroneous enumerations some people who should have been coded erroneous due to other residences.

1. Background

In past censuses, we have had no way to evaluate the A.C.E. coding of people duplicated outside the post-enumeration survey's search area. For Census 2000 there is an analysis project that searched for duplicates through out the country. Matching work was conducted to determine the extent of duplicate enumerations that were not found in the A.C.E. See Mule (2001) for more details. This allowed us to evaluate the A.C.E. coding of people duplicated outside the search area.

A person record found duplicated was captured twice by the census. The record of the person that was captured in the correct place according to census residence rules should have been coded as a correct enumeration. The record of the person that was captured in an incorrect place according to census residence rules should have been coded as an erroneous enumeration.

This analysis looks at how the A.C.E. coded E-Sample people duplicated to people outside the A.C.E. search area. We obtained large standard errors for many of the numbers of interest. This was due to the small number of cases considered in this study. This limited our ability to make meaningful interpretations.

1.1 Housing Unit Duplication Operation

This research into person record duplication across the country was brought about in part to evaluate the Housing Unit Duplication Operation. There was a two stage process to remove duplicate housing units and the people in them. The first stage identified 5.9 million person records as potential duplicates and flagged them for potential deletion. These person records were temporally removed from the census, they were also excluded from the E Sample. The second stage analyzed the potential duplicates to determine which ones were actual duplicates, 2.3 million person records were reinstated back into the census (they were still excluded from the E-Sample universe). The remaining person records were deleted from the census. For complete details, see Nash (2000).

1.2 Timing of the Enumeration

The timing of the various operations are shown in the following tables.

<u>Operation</u>	<u>Timing (Year 2000)</u>
Census housing unit	
Census mailout and update/leave delivery	March 3 - March 30
Non-response follow-up	April 7 - June 26
Group quarters	
Service-based enumeration (shelters, soup kitchens, outdoor locations)	March 27 - March 29
T-Night (household enumeration of RV parks, campgrounds, marinas, carnival workers, etc.)	March 31
All other group quarters	April 1 - May 6
A.C.E	
A.C.E. person interviewing	April 24 - September 11
A.C.E. person matching and follow-up	October 5 - December 5

2. Methods

This report used data from the Census Person Duplication (Mule, 2001). It considers only duplicate links where an E-Sample person record is the source. It is also limited to cases where the model weight is greater than 0.5, in other words we only consider duplicate links that we have confidence in. All numbers in this report are weighted with the final E-Sample person weight and the unbiased probability of duplication for an E-Sample person record. See Appendix C for the assignment of the unbiased probability of duplication. Standard errors were produced in VPLX using simple Jackknife.

3. Limits

This report does not separately examine the issue of movers. It does not measure separately the amount of duplication due to people moving during the time frame the enumeration took place.

Assumption: The duplicates identified by the Census Person Duplication Operation are correct. This report only considers cases where the Census Person Duplication team has confidence in the link. The report does not examine the quality of the Census Person Duplication Operation.

Assumption: there was a high quality group quarters enumeration. If this is not the case, the results regarding E-Sample people duplicated to group quarters need to be reexamined.

4. Results

This report deals with E-Sample people duplicated outside the surrounding blocks. For E-Sample people duplicated to housing units we expected about half to be erroneous enumerations. Because the A.C.E. was a random sample, one would expect that about half the time, the A.C.E. would capture the person record in the housing unit that they should have been counted in according to census residence rules. These person records should have been coded correctly enumerated (or matched) by A.C.E. The other half of the time, the A.C.E. should capture the person record in the housing unit that was incorrect according to census residence rules. These person records should have been coded erroneously enumerated by A.C.E.

Table 1 shows the number and percent of E-Sample people duplicated to people outside the surrounding blocks by their final match code category. Table 1 breaks down the target people based on type of unit they lived in: housing unit, group quarters, reinstated unit and deleted unit. Table 1 shows that 14.16 percent of the E-Sample people duplicated to housing units were erroneously enumerated. This is lower than the fifty percent we expected. Possible explanations for this difference include:

- The computer matching of duplicates outside the search area might be incorrect. We do not believe that this was the explanation, because we only looked at those cases that we had confidence in (those cases that had a high probability of being linked correctly).
- The A.C.E. did not do an good job of identifying people who should have been coded as erroneous enumerations due to other residence. The percent other residence was 1.4 in the 2000 A.C.E. and 2.3 in the 1990 PES (these percents reflect the redistribution of people with unresolved status). The results of the Measurement Error Reinterview also measure this phenomenon.

We expected the coding of E-Sample people duplicated in reinstated and deleted units to be similar to that of people duplicated in housing units. If the Housing Unit Duplication Operation favored geographically correct units, we expect the percent erroneous enumeration to be lower for people duplicated in reinstated and deleted units.

E-Sample people duplicated to people in group quarters outside the surrounding blocks are easier to interpret and will be the focus of the rest of this paper.

Table 1 E-Sample Duplicates to People Outside the Surrounding Blocks

Final Match Codex	E-Sample Eligible HU	GQ	Reinstate	Delete
M	1,298,084 (26,633)	238,095 (30,524)	275,548 (19,652)	122,073 (13,974)
CE	564,144 (20,115)	120,382 (9,625)	191,701 (17,964)	91,286 (12,243)
EE	307,138 (16,774)	162,024 (9,531)	106,920 (12,411)	51,303 (10,552)
Total	2,169,366 (49,926)	520,501 (37,069)	574,169 (30,959)	264,662 (22,117)
Percent				
Pct Match	59.84 (0.88)	45.74 (3.18)	47.99 (2.58)	46.12 (4.00)
Pct Corr. Enum.	26.00 (0.71)	23.13 (1.58)	33.39 (2.38)	34.49 (3.67)
Pct Err. Enum.	14.16 (0.65)	31.13 (2.56)	18.62 (1.90)	19.38 (3.50)

These number are weighted with the final E-Sample weight and the unbiased probability of duplication. They only include cases where the model probability of being a duplicate is greater than 0.5.

Table 2 shows the number of E-Sample people duplicated outside the surrounding blocks to people in group quarters. It shows the different group quarters types. It also indicates whether or not the people in the group quarters could claim usual home elsewhere (UHE). Assuming that the group quarters' enumeration was correct, the distinction between group quarters that could claim UHE and those that could not claim UHE is important because the A.C.E. person matching operation should have coded these cases as follows (Childers, 2001):

If the person lived in a GQ where they...	the E-Sample person should have been coded...
could claim UHE	correctly enumerated
could not claim UHE	erroneous enumerated

Table 2 shows that over half the duplicates to group quarters were duplicates to college dorms. Local jails account for 8.5 percent of the duplicates to group quarters. Some people counted in local jails and other temporary group quarters may have actually been usual residents of the sample housing unit. However according to A.C.E. person matching procedures, the person in a local jail should have been coded as an erroneous enumeration if they were in the local jail on April 1, 2000.

Table 2 E-Sample People Duplicated outside the A.C.E. search area to people in Group Quarters by type of Group Quarters

GQ Type	Number	Percent of GQ UHE OK?	
101 Federal detention Center	1,254	0.24	
102 Federal prison	1,119	0.21	
103 State prison	23,219	4.46	
104 Local jail	44,387	8.53	
105 Halfway Houses (Correctional)	663	0.13	
106 Military prison	9,995	1.92	
107 Other Prison	3,972	0.76	
202 Neglected/abused juvenile (long term)	1,897	0.36	
204 Emotionally disturbed kids (long term)	2,969	0.57	
205 Delinquent kids (long term)	8,168	1.57	
209 Other Juvenile institution	3,000	0.58	
301 Nursing home	56,471	10.85	
400 Drug/Alcohol Abuse treatment	3,533	0.68	
401 Military hospital	2,837	0.55	
402 Civilian hospital	3,107	0.60	
403 Hospices	308	0.06	
404 Mentally ill hospital	4,899	0.94	
405 Mentally handicapped hospital	789	0.15	
406 Inst. for Deaf	42	0.01	
407 Inst. for Blind	611	0.12	
408 Other physically handicap	5,455	1.05	
501 College Dorm	271,158	52.10	
601 Military Barracks	23,503	4.52	Y
701 Homeless shelter adult/family	8,660	1.66	
702 Children's shelter	491	0.09	
704 Soup kitchen	3,879	0.75	Y
705 Mobile food van	628	0.12	Y
706 TNSOLs	1,021	0.20	
801 Drug/Alcohol Abuse Group home	3,881	0.75	Y
802 Mentally ill Group home	2,738	0.53	Y
803 Mentally retarded Group home	1,822	0.35	Y
804 Physically Handicapped Group home	46	0.01	Y
805 Other Group home	6,390	1.23	Y
901 Agricultural worker's dorms	1,045	0.20	Y
902 Other worker dorms	891	0.17	Y
903 Job corps dorms	1,566	0.30	Y
904 Staff Dorms: Military Hospital/Prison	1,114	0.21	Y
906 Religious GQ	2,802	0.54	Y
908 Hostels, YM/WCAs, etc	2,540	0.49	Y
911 Protective Oversight	6,742	1.30	Y
Other	889	0.17	
Total GQ	520,501		

These number are weighted with the final E-Sample weight. They only include cases where the model probability of being a duplicate is greater than 0.5.

Table 3 splits people duplicated in group quarters into two categories: those that could claim UHE and those that could not claim to have a usual home elsewhere. Those that could not claim

UHE were split into two categories: those in college dorm and those not in college dorms. Table 3 shows the number of matches, correct enumerations and erroneous enumerations. Table 3 also show the percent matched, correctly enumerated and erroneously enumerated. Assuming the group quarters enumeration was perfect, those duplicated to group quarters that could not claim UHE should all be erroneously enumerated and those duplicated to group quarters that could claim UHE should all be correctly enumerated.

Table 3 E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters (Standard Errors)

Final Match Code	GQ		
	Could not Claim UHE		Could Claim UHE
	Not a Dorm	Dorm	
M	103,871 (8,015)	93,846 (29,051)	40,378 (4,853)
CE	54,565 (6,026)	54,055 (6,932)	11,761 (2,471)
EE	31,320 (3,773)	123,257 (8,573)	7,447 (1,693)
Total	189,756 (11,028)	271,158 (34,806)	59,586 (5,915)
Percent			
Pct Match	54.74 (2.78)	34.61 (7.43)	67.76 (4.21)
Pct Corr. Enum.	28.76 (2.58)	19.93 (2.04)	19.74 (3.78)
Pct Err. Enum.	16.51 (1.87)	45.46 (6.58)	12.50 (2.64)

These number are weighted with the final E-Sample weight. They only include cases where the model probability of being a duplicate is greater than 0.5.

Appendix A contains data by different variables in an attempt to better understand our results.

Appendix A contains data on the following variables:

- Person follow-up respondent (Tables 4.A and B)
- Age (Tables 5.A and B)
- Census respondent (Tables 6.A and B)
- Whole vs partial household duplicate (Tables 7A and B)
- Domain (Tables 8.A and B)
-

5. Summary

A.C.E. person matching did not correctly code many E-Sample records identified as duplicates in the Census Person Duplication Operation.

Possible explanations include:

- The instructions indicating who to include on the census questionnaire were not completely understood or the instructions may have been understood, but were ignored by the respondent. Example include college students, people in local jail and people in nursing homes.
- The respondent may not realize that a household member was enumerated elsewhere.
- Some group quarters' enumeration may have been done using administrative records that did not reflect the residents as of April 1, 2000.
- Some group quarters are temporary, such as local jails. Some people counted here may actually be usual residents of the sample housing unit.
- The computer matching of duplicates outside the search area might be incorrect. We do not believe that this was a large part of the explanation, because we only looked at those cases that we had confidence in (those cases that had a high probability of being linked correctly).
- The A.C.E. did not do an optimal job of identifying people who should have been coded as erroneous enumerations due to other residence. The percent other residence was 1.4 in the 2000 A.C.E. and 2.3 in the 1990 PES (these percentages reflect the redistribution of people with unresolved status). The results of the Measurement Error Reinterview also measure this phenomenon.

References

Childers, D. (January 2001), "Accuracy and Coverage Evaluation: The Design Document", DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-01.

Haines, D. (March, 2001), "Accuracy and Coverage Evaluation Survey: Computer Specifications for Person Dual System Estimation (U.S.) -Re-issue of Q-37", DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter Q-48.

Mule, T. (October 2001), "Executive Steering Committee on Accuracy and Coverage Evaluation Policy II Report Number 20: Person Duplication in Census 2000".

Nash F. (November 2000), "Overview of the Duplicate Housing Unit Operations".

Philipp, D. (February, 2001), "Hundred Percent Census Unedited File (HCUF)", 2000 Decennial Documentation.

Appendix A

Table 4.A E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters by PFU respondent

PFU/Final Match Code	GQ		
	Could Not Claim UHE		Could Claim UHE
	Not a Dorm	Dorm	
Non-proxy			
M	743 (630)	1,761 (951)	617 (507)
CE	30,639 (4,669)	44,358 (5,432)	6,485 (1,917)
EE	17,534 (2,804)	114,934 (8,323)	3,775 (1,244)
total	48,916 (5,571)	161,053 (10,033)	10,877 (2,349)
Proxy			
M	1,031 (737)	1,335 (1,237)	126 (126)
CE	23,927 (3,831)	9,697 (3,338)	5,276 (1,562)
EE	12,869 (2,483)	7,516 (1,813)	3,337 (1,092)
total	37,827 (4,700)	18,547 (4,988)	8,739 (1,945)
No Follow-up			
M	102,096 (7,929)	90,751 (27,234)	39,635 (4,743)
CE	0 (0)	0 (0)	0 (0)
EE	917 (523)	807 (531)	335 (265)
total	103,013 (7,960)	91,558 (27,243)	39,971 (4,831)

These number are weighted with the final E-Sample weight. They only include cases where the model probability of being a duplicate is greater than 0.5.

Table 4.B Percent Match, Correct Enumeration and Erroneous Enumeration of E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters by PFU Respondent

PFU/Final Match Code	GQ		
	Could Not Claim UHE		Could Claim UHE
	Not a Dorm	Dorm	
Non-Proxy			
Pct Match	1.52 (1.30)	1.09 (0.59)	5.67 (4.75)
Pct Corr. Enum.	62.64 (5.14)	27.54 (2.83)	59.62 (10.81)
Pct Err. Enum.	35.85 (5.05)	71.36 (2.93)	34.71 (10.34)
Proxy			
Pct Match	2.73 (1.95)	7.20 (6.58)	1.44 (1.48)
Pct Corr. Enum.	63.25 (5.71)	52.28 (8.01)	60.38 (10.91)
Pct Err. Enum.	34.02 (5.57)	40.52 (13.18)	38.18 (10.74)
No FU			
Pct Match	99.11 (0.51)	99.12 (0.68)	99.16 (0.66)
Pct Corr. Enum.	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Pct Err. Enum.	0.89 (0.51)	0.88 (0.68)	0.84 (0.66)

These rates are based on numbers weighted with the final E-Sample weight. They only include cases where the model probability of being a duplicate is greater than 0.5.

Table 5.A E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters by Age Group

Age/Final Match Code	GQ		
	Could Not Claim UHE		Could Claim UHE
	Not a Dorm	Dorm	
0-17			
M	13,375 (2,409)	2,204 (1,885)	4,756 (1,444)
CE	7,244 (1,784)	654 (654)	2,083 (1,036)
EE	2,995 (1,040)	978 (718)	676 (478)
total	23,614 (3,348)	3,836 (2,121)	7,516 (1,841)
18-23			
M	18,859 (3,156)	65,611 (7,299)	10,109 (2,490)
CE	7,006 (2,042)	49,025 (5,409)	4,419 (1,657)
EE	3,128 (1,117)	119,644 (8,407)	2,132 (979)
total	28,993 (3,957)	234,280 (12,278)	16,660 (3,160)
24-29			
M	14,553 (2,909)	15,321 (13,612)	4,941 (1,594)
CE	2,682 (1,213)	3,574 (3,574)	355 (311)
EE	3,181 (1,199)	2,605 (1,110)	602 (426)
total	20,417 (3,647)	21,501 (17,345)	5,898 (1,678)
30-49			
M	26,375 (3,773)	8,236 (8,029)	14,039 (2,568)
CE	15,615 (2,845)	726 (627)	1,792 (779)
EE	3,811 (1,231)	30 (26)	2,263 (965)
total	45,801	8,991	18,094

	(45,002)	(8,647)	(3,111)
50-64			
M	4,977 (1,450)	2,474 (1,473)	1,354 (1,064)
CE	5,375 (3,037)	76 (76)	587 (567)
EE	794 (466)	0 (0)	0 (0)
total	11,146 (3,396)	2,550 (1,475)	1,941 (1,205)
65+			
M	25,733 (3,500)	0 (0)	5,179 (1,647)
CE	16,643 (2,978)	0 (0)	2,525 (1,127)
EE	17,409 (2,926)	0 (0)	1,774 (754)
total	59,785 (5,519)	0 (0)	9,478 (2,133)

These number are weighted with the final E-Sample weight. They only include cases where the model probability of being a duplicate is greater than 0.5.

Table 5.B Percent Match, Correct Enumeration and Erroneous Enumeration of E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters by Age Group

Age/Final Match Code	GQ		
	Could Not Claim UHE		Could Claim UHE
	Not a Dorm	Dorm	
0-17			
Pct Match	56.64 (6.81)	57.46 (43.53)	63.28 (12.83)
Pct Corr. Enum.	30.68 (6.24)	17.06 (23.69)	27.72 (12.35)
Pct Err. Enum.	12.68 (4.04)	25.49 (29.46)	9.00 (6.48)
18-23			
Pct Match	65.05 (6.71)	28.01 (2.56)	60.68 (9.67)
Pct Corr. Enum.	24.16 (6.17)	20.93 (2.07)	26.52 (8.85)
Pct Err. Enum.	10.79 (3.75)	51.07 (2.68)	12.80 (5.79)
24-29			
Pct Match	71.28 (6.98)	71.26 (30.26)	83.78 (9.27)
Pct Corr. Enum.	13.14 (5.29)	16.62 (16.66)	6.01 (5.54)
Pct Err. Enum.	15.58 (5.73)	12.12 (46.85)	10.21 (7.51)
30-49			
Pct Match	57.59 (5.23)	91.60 (31.40)	77.59 (5.92)
Pct Corr. Enum.	34.09 (5.00)	8.07 (23.13)	9.90 (4.20)
Pct Err. Enum.	8.32 (2.63)	0.33 (8.27)	12.51 (4.74)
50-64			
Pct Match	44.65 (17.72)	97.02 (4.33)	69.75 (43.87)
Pct Corr. Enum.	48.22 (19.71)	2.98 (4.33)	30.25 (43.87)
Pct Err. Enum.	7.13	0.00	0.00

	(4.82)	(0.00)	(0.00)
65+			
Pct Match	43.04		54.64
	(4.50)		(11.84)
Pct Corr. Enum.	27.84		26.64
	(4.18)		(10.78)
Pct Err. Enum.	29.12		18.72
	(4.16)		(7.95)

These rates are based on numbers weighted with the final E-Sample weight. They only include cases where the model probability of being a duplicate is greater than 0.5.

Table 6.A E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters by Census respondent

Final Match Code	GQ		
	Could Not Claim UHE		Could Claim UHE
	Not a Dorm	Dorm	
Mail Return			
M	78,309 (6,663)	80,410 (23,578)	32,248 (4,364)
CE	39,576 (5,177)	40,965 (5,737)	9,567 (2,276)
EE	26,481 (3,460)	100,011 (7,809)	5,831 (1,520)
total	144,365 (9,300)	221,386 (28,646)	47,646 (5,316)
Enum. non-proxy			
M	24,599 (3,923)	9,732 (2,594)	7,670 (1,815)
CE	13,769 (2,821)	9,796 (2,295)	1,812 (910)
EE	3,654 (1,282)	22,817 (3,417)	1,607 (705)
total	42,022 (5,193)	42,344 (4,878)	11,090 (2,178)
Enum. proxy			
M	963 (724)	3,704 (3,704)	459 (325)
CE	1,221 (599)	3,294 (2,156)	382 (322)
EE	1,185 (808)	429 (410)	9 (9)
total	3,369 (1,248)	7,428 (4,810)	850 (464)

These number are weighted with the final E-Sample weight. They only include cases where the model probability of being a duplicate is greater than 0.5.

Table 6.B Percent Match, Correct Enumeration and Erroneous Enumeration of E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters by Census Respondent

Final Match Code	GQ		
	Could Not Claim UHE		Could Claim UHE
	Not a Dorm	Dorm	
Mail			
Pct Match	54.24 (3.15)	36.32 (7.14)	67.68 (4.82)
Pct Corr. Enum.	27.41 (2.95)	18.50 (1.94)	20.08 (4.34)
Pct Err. Enum.	18.34 (2.22)	45.17 (6.58)	12.24 (3.00)
Enum. non-proxy			
Pct Match	58.54 (5.85)	22.98 (5.37)	69.16 (9.17)
Pct Corr. Enum.	32.77 (5.44)	23.13 (4.84)	16.34 (7.82)
Pct Err. Enum.	8.70 (2.98)	53.88 (5.86)	14.49 (6.17)
Enum. proxy			
Pct Match	28.58 (21.26)	49.87 (53.30)	54.02 (41.39)
Pct Corr. Enum.	36.24 (18.84)	44.35 (47.32)	44.97 (40.28)
Pct Err. Enum.	35.18 (22.00)	5.78 (9.48)	1.01 (1.14)

These rates are based on numbers weighted with the final E-Sample weight. They only include cases where the model probability of being a duplicate is greater than 0.5.

Table 7.A E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters by Whole vs Partial HH Duplicate (Standard Errors)

Final Match Code	GQ		
	Could Not Claim UHE		Could Claim UHE
	Not a Dorm	Dorm	
Single Person HH			
M	16,347 (2,773)	1,104 (739)	7,717 (1,884)
CE	11,240 (2,282)	1,322 (676)	2,466 (1,122)
EE	9,135 (2,134)	524 (506)	1,443 (701)
total	36,721 (4,176)	2,950 (1,195)	11,626 (2,301)
Whole HH Duplicate			
M	7,484 (2,857)	27,258 (25,939)	3,437 (1,324)
CE	3,742 (1,607)	9,510 (4,412)	588 (383)
EE	3,042 (1,104)	3,984 (1,256)	1,495 (832)
total	14,269 (3,462)	40,752 (29,775)	5,520 (1,937)
Partial HH Duplicate			
M	80,040 (6,624)	65,484 (6,376)	29,224 (3,987)
CE	39,584 (5,350)	43,224 (4,828)	8,707 (2,170)
EE	19,143 (2,921)	118,748 (8,439)	4,510 (1,298)
total	138,767 (9,269)	227,456 (11,753)	42,440 (4,723)

These number are weighted with the final E-Sample weight. They only include cases where the model probability of being a duplicate is greater than 0.5.

Table 7.B Percent Match, Correct Enumeration and Erroneous Enumeration of E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters by Whole vs Partial HH Duplicate (Standard Errors)

Final Match Code	GQ		
	Could Not Claim UHE		Could Claim UHE
	Not a Dorm	Dorm	
Single Person HH			
Pct Match	44.52 (5.73)	37.43 (24.24)	66.38 (9.78)
Pct Corr. Enum.	30.61 (5.29)	44.80 (20.78)	21.21 (8.89)
Pct Err. Enum.	24.88 (5.07)	17.77 (18.32)	12.41 (6.02)
Whole HH Duplicate			
Pct Match	52.45 (13.48)	66.89 (55.16)	62.26 (12.26)
Pct Corr. Enum.	26.23 (11.03)	23.34 (29.99)	10.66 (7.36)
Pct Err. Enum.	21.32 (8.41)	9.78 (25.53)	27.08 (11.49)
Partial HH Duplicate			
Pct Match	57.68 (3.21)	28.79 (2.33)	68.86 (5.16)
Pct Corr. Enum.	28.53 (3.09)	19.00 (1.91)	20.52 (4.61)
Pct Err. Enum.	13.80 (1.99)	52.21 (2.61)	10.63 (2.99)

These percent are based on number are weighted with the final E-Sample weight. They only include cases where the model probability of being a duplicate is greater than 0.5.

Table 8.A E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters by Domain (Standard Errors)

	Could Not Claim UHE		Could Claim UHE
Domain/Final Match Code	Not a Dorm	Dorm	
American Indian on Res.			
M	403 (114)	87 (67)	97 (48)
CE	67 (39)	18 (18)	71 (41)
EE	135 (87)	51 (36)	0 (0)
total	605 (165)	156 (78)	167 (63)
American Indian off Res.			
M	1,795 (863)	0 (0)	0 (0)
CE	207 (112)	0 (0)	56 (56)
EE	590 (349)	84 (84)	0 (0)
total	2,593 (961)	84 (84)	56 (56)
Hispanic			
M	16,762 (3,135)	6,193 (2,469)	5,258 (1,430)
CE	4,706 (1,425)	4,024 (1,405)	680 (425)
EE	1,939 (829)	5,250 (1,635)	358 (330)
total	23,407 (3,604)	15,466 (3,276)	6,295 (1,538)
Black			
M	25,228 (3,528)	11,506 (2,262)	6,521 (1,658)
CE	15,297 (2,962)	8,799 (1,878)	2,385 (999)
EE	6,243 (1,509)	17,832 (3,069)	1,676 (733)
total	46,767 (5,136)	38,137 (4,267)	10,582 (2,075)
Pacific Islander			

M	126 (126)	126 (126)	0 (0)
CE	14 (14)	126 (126)	0 (0)
EE	0 (0)	0 (0)	0 (0)
total	140 (126)	251 (178)	0 (0)
Asian			
M	552 (393)	16,122 (10,609)	1,247 (726)
CE	433 (332)	3,663 (1,623)	179 (136)
EE	23 (23)	2,482 (1,049)	126 (126)
total	1,009 (515)	22,267 (11,927)	1,552 (750)
White			
M	59,006 (6,044)	59,813 (16,880)	27,255 (4,093)
CE	33,840 (5,050)	37,426 (5,637)	8,391 (2,216)
EE	22,389 (3,347)	97,559 (7,770)	5,288 (1,485)
total	115,236 (8,740)	194,798 (21,855)	40,934 (5,076)

These number are weighted with the final E-Sample weight. They only include cases where the model probability of being a duplicate is greater than 0.5.

Table 8.B Percent Match, Correct Enumeration and Erroneous Enumeration of E-Sample Duplicates to People Outside the A.C.E. Search Area in Group Quarters by Domain (Standard Errors)

Domain/Final Match Code	GQ		
	Could Not Claim UHE		Could Claim UHE
	Not a Dorm	Dorm	
American Indian on Res.			
Pct Match	66.59 (12.04)	55.97 (34.70)	57.78 (21.84)
Pct Corr. Enum.	11.07 (6.52)	11.56 (14.38)	42.22 (21.84)
Pct Err. Enum.	22.33 (12.75)	32.47 (29.30)	0.00 (0.00)
American Indian off Res.			
Pct Match	69.24 (17.79)	0.00 (0.00)	0.00 (0.00)
Pct Corr. Enum.	8.00 (4.96)	0.00 (0.00)	100.00 (0.00)
Pct Err. Enum.	22.76 (14.40)	100.00 (0.00)	0.00 (0.00)
Hispanic			
Pct Match	71.61 (6.62)	40.04 (12.04)	83.52 (8.88)
Pct Corr. Enum.	20.10 (5.69)	26.02 (8.81)	10.80 (6.83)
Pct Err. Enum.	8.29 (3.46)	33.94 (9.92)	5.68 (5.39)
Black			
Pct Match	53.94 (5.04)	30.17 (5.12)	61.63 (9.87)
Pct Corr. Enum.	32.71 (4.97)	23.07 (4.47)	22.54 (8.65)
Pct Err. Enum.	13.35 (3.03)	46.76 (5.66)	15.84 (6.72)
Pacific Islander			
Pct Match	89.72 (90.30)	50.00 (70.70)	
Pct Corr. Enum.	10.28 (90.30)	50.00 (70.70)	
Pct Err. Enum.	0.00 (0.00)	0.00 (0.00)	

Asian			
Pct Match	54.73 (35.26)	72.40 (19.76)	80.38 (17.82)
Pct Corr. Enum.	42.95 (35.17)	16.45 (8.07)	11.52 (11.99)
Pct Err. Enum.	2.31 (2.84)	11.15 (12.99)	8.10 (10.09)
White			
Pct Match	51.20 (3.72)	30.71 (6.25)	66.58 (5.39)
Pct Corr. Enum.	29.37 (3.53)	19.21 (2.21)	20.50 (4.91)
Pct Err. Enum.	19.43 (2.70)	50.08 (6.05)	12.92 (3.38)

Appendix B Technical Documentation

The following files were used in the creation of this report.

- Duplicate file
- Hundred-percent Census Unedited File (HCUF) (see Phillips (2001) for details)
- Keyed A.C.E. person follow-up data files (see Raglin et al (2001) for details)
- E-Sample Person Dual System Estimation (DSE) Output File (Haines, 2001)

Estimated E-Sample Components: Person records were given enumeration probabilities based on the following DSE file variables:

- < The person record's final match code (FINMAT).
- < The person's final probability of correct enumeration (CEPROBF). This includes the person's probability of being duplicated to a surrounding block

Type of Enumeration	Final Match codes	Enumeration probability
Matches:	FINMAT=M, MR	ceprobf
Correct Enumeration:	FINMAT=CE	ceprobf
Erroneous Enumeration:	all FINMAT	1-ceprobf

This report only considers duplicates out side of the surrounding block (clustgeo=3) and duplicate links we are confident in (mweight > 0.5). The numbers are weighted with the final E-sample person sampling weight (tesfinwt), the enumeration probability (defined above) and the unduplication probability (defined in Appendix C).

Variable	Variable Name	File	Collapsing
Housing unit type	etarget	Duplicate file	
GQ type	tqgqtyp	Duplicate file	Can claim UHE: tqgqtyp=601-604, 704, 705, 801-911 Can not claim UHE: tqgqtyp=100-500, 701-703, 706 Dorm: tqgqtyp=501
PFU respondent	resprel, fuflag	Keyed A.C.E. person follow-up, DSE	No follow-up: fuflag=blank, 1 Non-proxy: resprel=1, 02, 001, etc Proxy: otherwise
Age	sqage	Duplicate file	0-17, 18-23, 24-29, 30-49, 50-64, 65+
Response Method	pft, pcmode, rhhmem	HCUF	enumerator: pcmode = 2 or pft=5, 6, 17, 18 non-proxy: enumerator & rhhmem=0, 1 proxy: enumerator & rhhmem=2, 3
Whole/Partial HH duplicates		Duplicate file, DSE	The number of person records for each MAFID were counted on both files. Single: there was exactly one person record for the MAFID on the DSE file Whole HH duplicate: the number of person records on the duplicate file equals the number on the DSE file Partial HH duplicate: the number of person records on the duplicate file less than the number on the DSE file
Domain	domain	DSE	

Appendix C: Assignment of the Unbiased Probability of Duplication

For each duplication link between the Source and Target file identified by this analysis, we need to assign an unbiased probability of duplication. Our methodology is similar to Mule (2001). However there is one difference.

Since we examined person matching in A.C.E., our analysis used only E-Sample persons (E-sample Eligible persons selected for the E sample) in the clusters. Mule (2001) used all of the E-Sample Eligible persons in the cluster when assigning the probability. Our assignment reflects this difference.

Table C1: Combinations of Duplicates

Combination
Duplication of E-Sample Persons to E-Sample Eligible
Duplication of E-Sample Persons to GQ
Duplication of E-Sample Persons to Reinstate
Duplication of E-Sample Persons to Delete

Table C2 divides the records on the Source and Target files into 8 categories. The rest of this section describes how to assign probabilities based on the duplicate links between the categories.

Table C2: Categories for Assigning Unbiased Probabilities

Category	File	Description
A	Source and Target	E-sample Eligible People selected for the E sample in A.C.E. clusters (E-sample Persons)
B	Target	E-sample Eligible People not selected for the E sample in A.C.E. clusters and E-sample Eligible People not in A.C.E. clusters
C	Target	Group Quarters people in A.C.E. clusters
D	Target	Group Quarters people not in A.C.E. clusters
E	Target	Reinstated People in A.C.E. clusters
F	Target	Reinstated People not in A.C.E. clusters
G	Target	Deleted People in A.C.E. clusters
H	Target	Deleted People not in A.C.E. clusters

Table C3 shows how to assign the unbiased probabilities based on the category of each record. This table lists only the combinations for the estimates in this analysis.

Table C3: Assignment of Unbiased Duplication Probabilities

Source to Target Link			Duplication Probability Value
A	to	A	$\frac{1}{U + 1}$
A	to	B	$\left(\frac{1}{U + V + 1} \right) \frac{1}{U + 1}$
A	to	C or D	1
A	to	E or F	1
A	to	G or H	1

where U is the number of links from this Source A record to other category A records
V is the number of links from this Source A record to category B records.